Building the Segmented Linear Regression Model to Evaluate the Effectiveness of the Sterile Blood Culture Intervention.

We followed the template for segmented linear regression model building for interrupted time series data described by Wagner et al.[1] Time series data are measures of the same variable recorded in regular time intervals. In our study, the percentage of blood cultures contaminated among all cultures collected in the ED was computed in two-week time intervals. The percentage of cultures contaminated was the primary outcome and used as the dependent variable in the final model. A time series is said to be "interrupted" when there is a change in the process of interest that may result in changes to the pattern of data under surveillance. In our study, the regular pattern of blood culture contamination was interrupted with the implementation of the sterile blood culture intervention on January 31, 2010.

Our goal was to evaluate if implementation of this intervention resulted in a statistically significant change in the percentage of blood cultures contaminated. Segmented linear regression of interrupted time series data is one technique to make this evaluation. Implementation of our intervention created two time series segments: the baseline period containing 24 intervals (i.e. data points), and the intervention period containing 24 intervals. The implementation phase between the baseline and intervention periods (encompassing two data points) was considered a transition period and not included in the analysis. This transition period was identified *a priori* as the period to "work out the kinks" in the implementation process.

Linear regression was used to fit linear trend lines to each segment of our time series. Each segment was defined by two parameters—level (percentage of contaminated cultures) and linear trend (slope of the segment).

Similar to Table 2 in the Wagner et al[1] article, our data were formatted in a spreadsheet with 48 rows containing data for two-week intervals; model parameters included the percentage of cultures contaminated (study outcome), the interval number (I), the interval number since the implementation of the intervention (Ii), and the period (P). Period (P) was a binary variable indicating the baseline period or intervention period, and was the primary exposure variable of interest. Interval number (I) and interval number since implementation (Ii) were variables for the baseline period and intervention period secular trends, respectively, and were considered covariates in the model.

A multivariate linear regression model was built:

$$Y_t = B_0 + B_1 I_t + B_2 P_t + B_3 Ii_t + \text{error term}_t$$

Where,

Y_t = percentage of blood culture contaminated at interval *t*.

I = study time at interval *t*, numbered 1 through 48 starting with first point in baseline period and ending with last point in intervention period.

Ii = time since intervention at interval *t*, numbered 1 through 24 starting with first point in the intervention period and ending with last point in intervention period.

P = period = binary variable for interval *t* set to indicate the baseline period vs. intervention period.

B_0 = estimates baseline segment level

B_1 = estimates baseline segment trend

B_2 = estimates level change at time of intervention = immediate effect of intervention

B_3 = estimates intervention period trend

In this model, B_2 provided an estimate for the immediate effect of the intervention on the percentage of blood cultures contaminated after adjustment for the baseline period trend and intervention period trend.

We made two alterations to this basic model for our calculations. In order to stabilize variances in the model, the dependent variable (percentage of blood cultures contaminated) was log-transformed. Additionally, because of the time series nature of our data and the potential for autocorrelation between consecutive observations, we used an autoregressive integrated moving average (ARIMA (1, 0, 0)) model to correct for first-order autocorrelation—i.e. the tendency for error terms of consecutive data points in the model to be correlated due to their temporal relationship.

In STATA/IC 11.1 (College Station, TX), the ARIMA model was created using the "arima" command. The dependent variable was the log-transformed percentage of cultures contaminated. The independent variables included I, Ii, and P as outlined above.

Additional details on interrupted time-series analyses can be found in the excellent tutorial by Wagner et al.[1]

Reference

1. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002;27:299–309.